

03cc

OIPE

RAW SEQUENCE LISTING
 PATENT APPLICATION: US/09/851,410A

DATE: 12/04/2001
 TIME: 12:15:47

Input Set : N:\Crf3\RULE60\09851410A.txt
 Output Set: N:\CRF3\12042001\I851410A.raw

SEQUENCE LISTING

3 (1) GENERAL INFORMATION:
 5 (i) APPLICANT: Reyes, Gregory R
 6 Yarbough, Patrice O
 7 Bradley, Daniel W
 8 Krawczynski, Krzysztof Z
 9 Tam, Albert
 10 Fry, Kirk E
 12 (ii) TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
 13 Non-A/Non-B Hepatitis Viral Agent
 15 (iii) NUMBER OF SEQUENCES: 20
 17 (iv) CORRESPONDENCE ADDRESS:
 18 (A) ADDRESSEE: Dehlinger & Associates
 19 (B) STREET: 350 Cambridge Avenue, Suite 250
 20 (C) CITY: Palo Alto
 21 (D) STATE: CA
 22 (E) COUNTRY: USA
 23 (F) ZIP: 94306
 25 (v) COMPUTER READABLE FORM:
 26 (A) MEDIUM TYPE: Floppy disk
 27 (B) COMPUTER: IBM PC compatible
 28 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 29 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
 31 (vi) CURRENT APPLICATION DATA:
 32 (A) APPLICATION NUMBER: US/09/851,410A
 33 (B) FILING DATE: 07-May-2001
 35 (vii) PRIOR APPLICATION DATA:
 36 (A) APPLICATION NUMBER: 09/128,275
 37 (B) FILING DATE: 1998-08-03
 38 (A) APPLICATION NUMBER: US 07/681,078
 39 (B) FILING DATE: 05-APR-1991
 40 (A) APPLICATION NUMBER: US 07/505,888
 41 (B) FILING DATE: 05-APR-1990
 42 (A) APPLICATION NUMBER: US 07/420,921
 43 (B) FILING DATE: 13-OCT-1989
 44 (A) APPLICATION NUMBER: US 07/367,486
 45 (B) FILING DATE: 16-JUN-1989
 46 (A) APPLICATION NUMBER: US 07/336,672
 47 (B) FILING DATE: 11-APR-1989
 48 (A) APPLICATION NUMBER: US 07/208,997
 49 (B) FILING DATE: 17-JUN-1988
 50 (viii) ATTORNEY/AGENT INFORMATION:
 51 (A) NAME: Petithory, Joanne R.
 52 (B) REGISTRATION NUMBER: 42,995
 53 (C) REFERENCE/DOCKET NUMBER: 4600-0183.24
 54 (ix) TELECOMMUNICATION INFORMATION:
 55 (A) TELEPHONE: (650) 324-0880

ENTERED

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/851,410A

DATE: 12/04/2001
TIME: 12:15:47

Input Set : N:\Crf3\RULE60\09851410A.txt
Output Set: N:\CRF3\12042001\I851410A.raw

70 (B) TELEFAX: (650) 324-0960
72 (2) INFORMATION FOR SEQ ID NO: 1:
74 (i) SEQUENCE CHARACTERISTICS:
75 (A) LENGTH: 1295 base pairs
76 (B) TYPE: nucleic acid
77 (C) STRANDEDNESS: double
78 (D) TOPOLOGY: linear
W--> 80 (ii) MOLECULE TYPE: DNA
82 (iii) HYPOTHETICAL: NO
84 (iv) ANTI-SENSE: NO
86 (vi) ORIGINAL SOURCE:
87 (C) INDIVIDUAL ISOLATE: 1.33 kb EcoRI insert of ET1.1,
88 forward sequence
90 (ix) FEATURE:
91 (A) NAME/KEY: CDS
92 (B) LOCATION: 1..1293
94 (ix) FEATURE:
95 (A) NAME/KEY: CDS
96 (B) LOCATION: 2..1294
98 (ix) FEATURE:
99 (A) NAME/KEY: CDS
100 (B) LOCATION: 3..1295
103 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
105 AGACCTGTCC CTGTTGCAGC TGTTCTACCA CCCTGCCCCG AGCTCGAACCA GGGCCTTCTC 60
107 TACCTGCCCG AGGAGCTCAC CACCTGTGAT AGTGTGTTAA CATTGAAATT AACAGACATT 120
109 GTGCACTGCC GCATGGCCGC CCCGAGCCAG CGCAAGGCCG TGCTGTCCAC ACTCGTGGGC 180
111 CGCTACGGCG GTCGACACAA GCTCTACAAT GCTTCCCCTTG ATGATGTTCG CGACTCTCTC 240
113 GCCCGTTTTA TCCCGGCCAT TGGCCCCGTA CAGGTTACAA CTTGTGAATT GTACGAGCTA 300
115 GTGGAGGCCA TGGTCGAGAA GGGCCAGGAT GGCTCCGCCG TCCTTGAGCT TGATCTTTGC 360
117 AACCGTGACG TGTCCAGGAT CACCTTCTTC CAGAAAGATT GTAACAAGTT CACCACAGGT 420
119 GAGACCATTG CCCATGGTAA AGTGGGCCAG GGCATCTCGG CCTGGAGCAA GACCTTCTGC 480
121 GCCCTCTTTG GCCCTTGGTT CCGCGCTATT GAGAAGGCTA TTCTGGCCCT GCTCCCTCAG 540
123 GGTGTGTTTT ACGGTGATGC CTTTGATGAC ACCGTCTTCTC CGCGGGCTGT GGCCGCAGCA 600
125 AAGGCATCCA TGGTGTGTTGA GAATGACTTT TCTGAGTTTG ACTCCACCCA GAATAACTTT 660
127 TCTCTGGGTG TAGAGTGTGC TATTATGGAG GAGTGTGGGA TGCCGCAGTG GCTCATCCGC 720
129 CTGTATCACC TTATAAGGTC TGCGTGGATC TTGCGAGGCC CGAAGGAGTC TCTCGAGGG 780
131 TTTTGGAAAGA AACACTCCGG TGAGCCCGC ACTCTTCTAT GGAATACTGT CTGGAATATG 840
133 GCCGTTATTA CCCACTGTTA TGACTTCCGC GATTTCAGG TGGCTGCCTT TAAAGGTGAT 900
135 GATTGATAG TGCTTGCAG TGAGTATCGT CAGAGTCCAG GAGCTGCTGT CCTGATGCC 960
137 GGCTGTGGCT TGAAGTTGAA GGTAGATTC CGCCCGATCG GTTGTATGC AGGTGTTGTG 1020
139 GTGGCCCCCG GCCTTGGCGC GCTCCCTGAT GTTGTGCGCT TCGCCGGCCG GCTTACCGAG 1080
141 AAGAATTGGG GCCCTGGCCC TGAGCGGGCG GAGCAGCTCC GCCTCGCTGT TAGTGATTTC 1140
143 CTCCGCAAGC TCACCGAATGT AGCTCAGATG TGTGTGGATG TTGTTTCCCG TGTTTATGGG 1200
145 GTTTCCCCCTG GACTCGTTCA TAACCTGATT GGCATGCTAC AGGCTGTTGC TGATGGCAAG 1260
147 GCACATTTCA CTGAGTCAGT AAAACCAGTG CTCGA 1295
150 (2) INFORMATION FOR SEQ ID NO: 2:
152 (i) SEQUENCE CHARACTERISTICS:
153 (A) LENGTH: 431 amino acids
154 (B) TYPE: amino acid

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/851,410A

DATE: 12/04/2001
TIME: 12:15:47

Input Set : N:\Crf3\RULE60\09851410A.txt
Output Set: N:\CRF3\12042001\I851410A.raw

155 (D) TOPOLOGY: linear
157 (ii) MOLECULE TYPE: protein
159 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
161 Arg Pro Val Pro Val Ala Ala Val Leu Pro Pro Cys Pro Glu Leu Glu
162 1 5 10 15
164 Gln Gly Leu Leu Tyr Leu Pro Gln Glu Leu Thr Thr Cys Asp Ser Val
165 20 25 30
167 Val Thr Phe Glu Leu Thr Asp Ile Val His Cys Arg Met Ala Ala Pro
168 35 40 45
170 Ser Gln Arg Lys Ala Val Leu Ser Thr Leu Val Gly Arg Tyr Gly Gly
171 50 55 60
173 Arg Thr Lys Leu Tyr Asn Ala Ser His Ser Asp Val Arg Asp Ser Leu
174 65 70 75 80
176 Ala Arg Phe Ile Pro Ala Ile Gly Pro Val Gln Val Thr Thr Cys Glu
177 85 90 95
179 Leu Tyr Glu Leu Val Glu Ala Met Val Glu Lys Gly Gln Asp Gly Ser
180 100 105 110
182 Ala Val Leu Glu Leu Asp Leu Cys Asn Arg Asp Val Ser Arg Ile Thr
183 115 120 125
185 Phe Phe Gln Lys Asp Cys Asn Lys Phe Thr Thr Gly Glu Thr Ile Ala
186 130 135 140
188 His Gly Lys Val Gly Gln Gly Ile Ser Ala Trp Ser Lys Thr Phe Cys
189 145 150 155 160
191 Ala Leu Phe Gly Pro Trp Phe Arg Ala Ile Glu Lys Ala Ile Leu Ala
192 165 170 175
194 Leu Leu Pro Gln Gly Val Phe Tyr Gly Asp Ala Phe Asp Asp Thr Val
195 180 185 190
197 Phe Ser Ala Ala Val Ala Ala Lys Ala Ser Met Val Phe Glu Asn
198 195 200 205
200 Asp Phe Ser Glu Phe Asp Ser Thr Gln Asn Asn Phe Ser Leu Gly Leu
201 210 215 220
203 Glu Cys Ala Ile Met Glu Glu Cys Gly Met Pro Gln Trp Leu Ile Arg
204 225 230 235 240
206 Leu Tyr His Leu Ile Arg Ser Ala Trp Ile Leu Gln Ala Pro Lys Glu
207 245 250 255
209 Ser Leu Arg Gly Phe Trp Lys Lys His Ser Gly Glu Pro Gly Thr Leu
210 260 265 270
212 Leu Trp Asn Thr Val Trp Asn Met Ala Val Ile Thr His Cys Tyr Asp
213 275 280 285
215 Phe Arg Asp Phe Gln Val Ala Ala Phe Lys Gly Asp Asp Ser Ile Val
216 290 295 300
218 Leu Cys Ser Glu Tyr Arg Gln Ser Pro Gly Ala Ala Val Leu Ile Ala
219 305 310 315 320
221 Gly Cys Gly Leu Lys Leu Lys Val Asp Phe Arg Pro Ile Gly Leu Tyr
222 325 330 335
224 Ala Gly Val Val Val Ala Pro Gly Leu Gly Ala Leu Pro Asp Val Val
225 340 345 350
227 Arg Phe Ala Gly Arg Leu Thr Glu Lys Asn Trp Gly Pro Gly Pro Glu
228 355 360 365

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/851,410A

DATE: 12/04/2001
TIME: 12:15:47

Input Set : N:\Crf3\RULE60\09851410A.txt
Output Set: N:\CRF3\12042001\I851410A.raw

230 Arg Ala Glu Gln Leu Arg Leu Ala Val Ser Asp Phe Leu Arg Lys Leu
231 370 375 380
233 Thr Asn Val Ala Gln Met Cys Val Asp Val Val Ser Arg Val Tyr Gly
234 385 390 395 400
236 Val Ser Pro Gly Leu Val His Asn Leu Ile Gly Met Leu Gln Ala Val
237 405 410 415
239 Ala Asp Gly Lys Ala His Phe Thr Glu Ser Val Lys Pro Val Leu
240 420 425 430

243 (2) INFORMATION FOR SEQ ID NO: 3:
245 (i) SEQUENCE CHARACTERISTICS:
246 (A) LENGTH: 18 base pairs
247 (B) TYPE: nucleic acid
248 (C) STRANDEDNESS: single
249 (D) TOPOLOGY: linear

W--> 251 (ii) MOLECULE TYPE: DNA
253 (iii) HYPOTHETICAL: NO
255 (iv) ANTI-SENSE: NO
257 (vi) ORIGINAL SOURCE:
258 (C) INDIVIDUAL ISOLATE: linker - top (5') sequence
260 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

18

262 GGAATTCGCG GCCGCTCG
265 (2) INFORMATION FOR SEQ ID NO: 4:
267 (i) SEQUENCE CHARACTERISTICS:
268 (A) LENGTH: 20 base pairs
269 (B) TYPE: nucleic acid
270 (C) STRANDEDNESS: single
271 (D) TOPOLOGY: linear

W--> 273 (ii) MOLECULE TYPE: DNA
275 (iii) HYPOTHETICAL: NO
277 (iv) ANTI-SENSE: NO
279 (vi) ORIGINAL SOURCE:
280 (C) INDIVIDUAL ISOLATE: linker - bottom (3') sequence
282 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

20

284 CGAGCGGCCG CGAATTCCCTT
286 (2) INFORMATION FOR SEQ ID NO: 5:
288 (i) SEQUENCE CHARACTERISTICS:
289 (A) LENGTH: 1295 base pairs
290 (B) TYPE: nucleic acid
291 (C) STRANDEDNESS: double
292 (D) TOPOLOGY: linear

W--> 294 (ii) MOLECULE TYPE: DNA
296 (iii) HYPOTHETICAL: NO
298 (iv) ANTI-SENSE: NO
300 (vi) ORIGINAL SOURCE:
301 (C) INDIVIDUAL ISOLATE: 1.33 kb EcoRI insert of ET1.1,
302 reverse sequence

305 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
307 TCGAGCACTG GTTTTACTGA CTCAGTGAAA TGTGCCTTGC CATCAGCAAC AGCCTGTAGC
309 ATGCCAATCA GGTTATGAAC GAGTCCAGGG GAAACCCCCAT AAACACGGGA AACAAACATCC

60

120

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/851,410A

DATE: 12/04/2001

TIME: 12:15:47

Input Set : N:\CrF3\RULE60\09851410A.txt

Output Set: N:\CRF3\12042001\I851410A.raw

311	ACACACATCT	GAGCTACATT	CGTGAGCTTG	CGGAGGAAAT	CACTAACAGC	GAGGCCGGAGC	180
313	TGCTCCGCC	GCTCAGGGCC	AGGGCCCCAA	TTCTTCTCGG	TAAGCCGGCC	GGCGAAGGCC	240
315	ACAACATCAG	GGAGCGCGCC	AAGGCCGGGG	GCCACCAAA	CACCTGCATA	CAAACCGATC	300
317	GGGCCGGAAAT	CTACCTTCAA	CTTCAAGCCA	CAGCGGGCGA	TCAGGACAGC	AGCTCCTGGA	360
319	CTCTGACGAT	ACTCACTGCA	AAGCACTATC	GAATCATCAC	CTTTAAAGGC	AGCCACCTGA	420
321	AAATCGCGGA	AGTCATAACA	GTGGGTAATA	ACGGCCATAT	TCCAGACAGT	ATTCCATAGA	480
323	AGAGTGCAGG	GCTCACCGGA	GTGTTTCTTC	CAAACCCCTC	GCAGAGACTC	CTTCGGGGCC	540
325	TGCAAGATCC	ACGCAGACCT	TATAAGGTGA	TACAGGCGGA	TGAGCCACTG	CGGCATCCCA	600
327	CACTCCTCCA	TAATAGCACA	CTCTAGACCC	AGAGAAAAGT	TATTCTGGGT	GGAGTCAAAC	660
329	TCAGAAAAAGT	CATTCTAAA	CACCATGGAT	GCCTTGCTG	CGGCCACAGC	CGCCGAGAAG	720
331	ACGGTGTAT	CAAAGGCATC	ACCGTAAAAC	ACACCCCTGAG	GGAGCAGGGC	CAGAATAGCC	780
333	TTCTCAATAG	CGCGGAACCA	AGGGCCAAAG	AGGGCGCAGA	AGGTCTTGCT	CCAGGCCGAG	840
335	ATGCCCTGGC	CCACTTTTAC	ATGGGCAATG	GTCTCACCTG	TGGTGAACCT	GTTACAATCT	900
337	TTCTGGAAGA	AGGTGATCCT	GGACACGTCA	CGGTTGCAAA	GATCAAGCTC	AAGGACGGCG	960
339	GAGCCATCCT	GGCCCTTCTC	GACCATGGCC	TCCACTAGCT	CGTACAATTC	ACAAGTTGTA	1020
341	ACCTGTACGG	GGCCAATGGC	CGGGATAAAA	CGGGCGAGAG	AGTCGCGAAC	ATCAGAGTGG	1080
343	GAAGCATTGT	AGAGCTTTGT	GCGACCGCCG	TAGCGGCCCA	CGAGTGTGGA	CAGCACGGCC	1140
345	TTGCGCTGGC	TCGGGGCGGC	CATGCGGCAG	TGCACAATGT	CTGTTAATTC	AAATGTTACG	1200
347	ACACTATCAC	AGGTGGTGAG	CTCCTGGGGC	AGGTAGAGAA	GGCCCTGTTC	GAGCTCGGGG	1260
349	CAGGGTGGTA	GAACAGCTGC	AACAGGGACA	GGTCT			1295

352 (2) INFORMATION FOR SEQ ID NO: 6:

354 (i) SEQUENCE CHARACTERISTICS:

355 (A) LENGTH: 7195 base pairs

356 (B) TYPE: nucleic acid

W--> 367 (C) STRANDEDNESS: HEV - Burma strain

358 (D) TOPOLOGY: linear

W--> 360 (ii) MOLECULE TYPE: DNA

362 (iii) HYPOTHETICAL: NO

364 (iv) ANTI-SENSE: NO

366 (vi) ORIGINAL SOURCE:

369 (ix) FEATURE:

370 (A) NAME/KEY: CDS

371 (B) LOCATION: 28..5106

373 (ix) FEATURE:

374 (A) NAME/KEY: CDS

375 (B) LOCATION: 5147..7126

377 (ix) FEATURE:

378 (A) NAME/KEY: CDS

379 (B) LOCATION: 5106..5474

382 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

384	AGGCAGACCA	CATATGTGGT	CGATGCCATG	GAGGCCCATC	AGTTTATTAA	GGCTCCTGGC	60
386	ATCACTACTG	CTATTGAGCA	GGCTGCTCTA	GCAGCGGCCA	ACTCTGCCT	GGCGAATGCT	120
388	GTGGTAGTTA	GGCCTTTCT	CTCTCACCAAG	CAGATTGAGA	TCCTCATCAA	CCTAATGCAA	180
390	CCTCGCCAGC	TTGTTTCCG	CCCCGAGGTT	TTCTGGAATC	ATCCCATCCA	GCGTGTATC	240
392	CATAACGAGC	TGGAGCTTTA	CTGCCCGGCC	CGCTCCGGCC	GCTGTCTGTA	AATTGGCGCC	300
394	CATCCCCGCT	CAATAAAATGA	TAATCCTAAT	GTGGTCCACC	GCTGCTTCCT	CCGCCCTGTT	360
396	GGCGGTGATG	TTCAGCGCTG	GTATACTGCT	CCCACTCGCG	GGCGGGCTGC	TAATTGCCGG	420
398	CGTCCCGCCG	TGCGGGGGCT	TCCCGCTGCT	GACCGCACTT	ACTGCTCTCGA	CGGGTTTTCT	480
400	GGCTGTAACT	TTCCCGCCGA	GAATGGCATC	GCCTCTACT	CCCTTCATGA	TATGTACCCA	540

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/851,410A

DATE: 12/04/2001

TIME: 12:15:48

Input Set : N:\Crf3\RULE60\09851410A.txt
Output Set: N:\CRF3\12042001\I851410A.raw

L:32 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:33 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:80 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=1
L:251 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=3
L:273 M:246 W: Invalid value of Alpha Sequence Header Field, [MCLECULE TYPE:], SeqNo=4
L:294 M:246 W: Invalid value of Alpha Sequence Header Field, [MCLECULE TYPE:], SeqNo=5
L:367 M:220 C: Keyword misspelled or invalid format, [(C) STRANDEDNESS:]
L:367 M:246 W: Invalid value of Alpha Sequence Header Field, [STRANDEDNESS:], SeqNo=6
L:360 M:246 W: Invalid value of Alpha Sequence Header Field, [MCLECULE TYPE:], SeqNo=6
L:1142 M:220 C: Keyword misspelled or invalid format, [(C) STRANDEDNESS:]
L:1142 M:246 W: Invalid value of Alpha Sequence Header Field, [STRANDEDNESS:], SeqNo=10
L:1135 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=10
L:1403 M:220 C: Keyword misspelled or invalid format, [(C) STRANDEDNESS:]
L:1403 M:246 W: Invalid value of Alpha Sequence Header Field, [STRANDEDNESS:], SeqNo=11
L:1396 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=11
L:1477 M:220 C: Keyword misspelled or invalid format, [(C) STRANDEDNESS:]
L:1477 M:246 W: Invalid value of Alpha Sequence Header Field, [STRANDEDNESS:], SeqNo=12
L:1470 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=12
L:1651 M:220 C: Keyword misspelled or invalid format, [(C) STRANDEDNESS:]
L:1651 M:246 W: Invalid value of Alpha Sequence Header Field, [STRANDEDNESS:], SeqNo=17
L:1680 M:220 C: Keyword misspelled or invalid format, [(C) STRANDEDNESS:]
L:1680 M:246 W: Invalid value of Alpha Sequence Header Field, [STRANDEDNESS:], SeqNo=18
L:1708 M:220 C: Keyword misspelled or invalid format, [(C) STRANDEDNESS:]
L:1708 M:246 W: Invalid value of Alpha Sequence Header Field, [STRANDEDNESS:], SeqNo=19
L:1736 M:220 C: Keyword misspelled or invalid format, [(C) STRANDEDNESS:]
L:1736 M:246 W: Invalid value of Alpha Sequence Header Field, [STRANDEDNESS:], SeqNo=20